

Non-representational Interaction Design

Marco Gillies¹ and Andrea Kleinsmith²

¹ Embodied Audio-Visual Interaction, Department of Computing
Goldsmiths, University of London, London, UK SE14 6NW

m.gillies@gold.ac.uk

<http://www.doc.gold.ac.uk/~mas02mg/>

² Virtual Experiences Research Group

Department of Computer and Information Science and Engineering
University of Florida

Abstract. This paper presents how non-representational views of cognition can inform interaction design as it moves from traditional graphical user interfaces to more bodily forms of interaction such as gesture or movement tracking. We argue that the true value of these “bodily” interfaces is that they can tap our prior skills for interacting in the world. However, these skills are highly non-representational and so traditional representational approaches to interaction design will fail to capture them effectively. We propose interactive machine learning as an alternative approach to interaction design that is able to capture non-representational sensori-motor couplings by allowing people to design by performing actions rather than by representing them. We present an example of this approach applied to designing interactions with video game characters.

Keywords: interaction design, bodily interaction, interactive machine learning.

Interaction Design is a discipline saturated with representations. The primary interaction mechanism with modern computers is a graphical interface which is composed largely of visual representations of the computer system. These can include a wide variety of displays representing anything from documents to scientific data sets to social interactions. The interface also includes many different representations of possible actions that people can take from simple buttons to directly manipulable visualisations. The implementations of the interfaces also require complex logical representations, as any feature of the world or computer system has to be represented in program code, normally with explicit representations of items in the world and explicit mapping of these representations to the visual interface.

These representations are in no way incompatible with a non-representational view of cognition. They are, after all, external representations, part of our distributed cognitive apparatus that we can manipulate to achieve our aims. They largely exist in domains where we have very well established forms of external representation that most educated people are well versed in manipulating as

part of our working process: written language, diagrams and photographs. In fact, the graphical user interface is now so ubiquitous that it is a key part of the distributed cognitive apparatus of most adults in the developed world.

So we could be happy that a representational approach to interaction design and a non-representational approach to cognition are compatible if it were not for certain new developments in interaction design. New developments in human computer interaction are leaving the confines of the computer monitor and entering the 3 dimensional world of our full body movements. This paper will argue that this move will force interaction design to take account of non-representational cognition.

1 Bodily Interaction

The use of body movement to interact with computers has a long history, going back to pioneering work from the 1970s by researchers such as Myron Krueger[1]. However, only in recent years has it become feasible to create low cost, mass market bodily interfaces, due to the advent of commercial movement tracking devices such as the Microsoft Kinect™ or the accelerometers built into modern smart phones. These new devices enable us to make use of large scale movements of many different parts of the body, in contrast to the small scale movements required by a keyboard and mouse. Much of the research in this areas has concerned gesture recognition and it's use in user interfaces, for example the work of Bevilacqua et al.[2] or Fails and Olsen[3]. However, there is also research that makes use of holistic movement in interaction ranging from dance controlled music (Antle et al. [4]) and expressive digital musical instruments (Fiebrink [5]) to interactive art (Snibbe [6]) and body monitoring for healthcare (Fergus [7]). With devices such as the Microsoft Kinect™ body movement interfaces are now being used by ordinary consumers, with the first area of growth being in video games, given that most devices are marketed as video game controllers. However, their use is spreading, even if they remain attached to video game consoles, body movement interfaces are being used for other activities, for example the WiiFit software uses a games console as a platform for exercise and health.

2 What Is Natural about “Natural User Interfaces”?

What is the value of this for interaction? The type of interaction I have been describing has been marketed by Microsoft and others as “Natural User Interfaces”: interfaces that are claimed to be so natural that they do not need to be learned. The logic behind this phrase is that, because body movements come naturally to us, a body movement interface will be natural. This idea has been criticised by many people, most notably by Norman in his article “Natural User Interfaces are not natural” [8] in which he argues that bodily interfaces can suffer from many problems associated with traditional interfaces (such as the difficulty of remembering gesture) as well as new problems (the ephemerality of gestures

and lack of visual feedback). So is there value in the intuition that bodily interfaces are natural, and if so what is that value and why is it often not seen in existing interfaces?

I would argue that there is a fundamental difference in the nature of bodily interfaces and traditional interfaces. Jacob et al.[9] propose that a variety of new forms of interaction, including bodily interaction, are successful because they leverage a different set of our pre-existing skills from traditional GUIs. While a graphical user interface leverages our skills in manipulating external visual and symbolic representations, bodily interfaces leverage skills related to body and environmental awareness. The skills that enable us to move and act in the world. Similarly, Dourish [10] proposes that we analyse interaction in terms of embodiment which he defines as: “the property of our engagement with the world that allows us to make it meaningful”. This leads him to define Embodied Interaction as “the creation, manipulation, and sharing of meaning through engaged interaction with artefacts”. While he applies this definition to both traditional and new forms of interaction, the nature of this engaged interaction is very different in bodily interfaces. Following Jacob we could say that, in a successful bodily interface, this engaged interaction can be the same form of engagement we have with our bodies and environment in our daily lives and we can therefore re-use our existing skills that enable us to engage with the world.

If we take a non-representational, sensorimotor view of perception and action these skills are very different from the skills of a traditional interface involving manipulation of representations. This view allows us to keep the intuition that bodily interfaces are different from graphical user interfaces and explain what is meant by natural in the phrase “natural user interface” (the so-called natural skills are non-representational sensorimotor skills), while also allowing us to be critical of the claims of bodily interfaces. Natural user interfaces, on this view, are only natural if they take account of the non-representational, sensorimotor nature of our body movement skills. Body movement interfaces which are just extensions of a symbolic, representational interface which are just a more physically tiring version of a GUI.

A good example of this is gestural interaction. A common implementation of this form of interface is to have a number of pre-defined gestures that can be mapped to actions in the interface. This is one of the types of interface that Norman[8] criticises. When done badly there is a fairly arbitrary mapping between a symbolic gesture and a symbolic action. Users’ body movements are used as part of a representation manipulation task. There is nothing wrong with this *per se* but it does not live up to the hype of natural user interfaces and is not much different from a traditional GUI. In fact, as Norman notes, it can be worse, as users do not have a visual cue to remind them which gestures they should be performing. This makes it closer to a textual command line interface where users must remember obscure commands with no visual prompts. Gestural user interfaces do not have to be like this.

These problems can be avoided if we think of gestural interfaces as tapping sensorimotor skills, not representation manipulation skills. For example, the work of Bevilacqua et al.[2] uses gesture to control music. In this work, gestures are tracked continuously rather than being simply recognised at the end of the gesture. This allows users to continuously control the production of sound throughout the time they are performing the gesture, rather than triggering the gesture at the end. This seemingly simple difference transforms the task from representation manipulation (producing a symbolic gesture and expecting a discrete response) to a tight sensorimotor loop in which the auditory feedback can influence movement which in turn controls the audio. A more familiar example of this form of continuous feedback is the touch screen “pinch to zoom” gesture developed for the iPhoneTM. In this gesture an image resizes dynamically and continuously in response to the users’ fingers moving together and apart. This continuous feedback and interaction enables a sensorimotor loop that can leverage our real world movement skills.

A second feature of Bevilacqua et al.’s[2] system is that it allows users to easily define their own gestures and they do so by acting out those gestures while listening to the music to be controlled. I will come back to this feature in more detail later, but for now we can note that it means that gestures are not limited to a set of pre-defined symbolic gestures. Users can define movements that feel natural to them for controlling a particular type of music. What does “natural” mean in this context? Again, it means that the user already has a learnt sensorimotor mapping between the music and a movement (for example a certain way of tapping the hands in response to a beat).

3 How Do We Design Non-representational Interactions?

This brings us to the question of how we design bodily interactions. Interaction design for traditional user interfaces is largely a task of designing representations, for example, the layout of widget or the display of information. If we take successful bodily interaction to require non-representational sensorimotor skills, then the task of designing it must be something very different. Rather than designing representations, it means designing sensorimotor couplings. In fact, designing sensori-motor couplings are not restricted to bodily interactions. The two are not as distinct as I have presented them here. Viewed from this perspective, direct manipulation GUIs involve elements of sensorimotor coupling and most bodily interfaces are likely to involve some external representations.

The key challenge is that, if we are designing sensorimotor couplings that we perform without mental or external representations, then we must design them without representations. If we do not use mental representations to perform these skills and we do not have a well established system of external representations to handle them, we are unlikely to be able to explicitly represent details of our movement skills. If we use traditional, representational methods of

interaction design we will be asking designers to provide explicit representations of low level details of their movements, to which they have no conscious access. This is likely to be an impossible task. This is one explanation of the problems of existing bodily interaction techniques. Designers have to explicitly represent their movements in numerical terms programme code or other formal tools but they do not have this level of access to complex sensorimotor skills. So instead, they fall back on representing simple features of the movement (responding to position or velocity) or falling back on traditional user interface metaphors (buttons in 3D space). Neither of these approaches are designed to make effective use of sensorimotor skills, rather they are driven by what can easily be represented in code. The result is interactions that are not “natural” in the terms we have discussed above and which are hard to remember and perform.

We need a different approach. If we are to successfully create complex movement interfaces we must develop software tools that support creating based on our existing, non-representational sensorimotor skills. That means we cannot ask designers to form explicit representations of their movements. Instead we should allow designers to design interactions by directly applying their bodily skills. They should define movements by moving.

Designing by moving can be enabled by software tools that allow designers to specify interaction by giving examples of those interactions. This type of tool can be implemented using machine learning techniques to infer recognition models from these examples. However, traditional batch approaches to machine learning are not particularly well suited to a design process as designers must collect large amounts of data initially and then must rely on the algorithm to produce the desired results based on this data. This makes it hard to support the iterative processes and refinement that are key to successful interaction design. Interactive Machine Learning (Fails and Olsen [3]) is an approach to machine learning which attempts to overcome these issues by making user interaction central to the learning process. Users provide training data interactively and in doing so progressively refine the classifiers they are creating. This approach has been used in a variety of domains from image classification (Talbot et al. [11]) to end user training of spam filters (Kulesza et al. [12]). This method can allow designers to design interactions by interactively providing examples of movement which would be used to train and refine a machine learning algorithm that controls the interaction. This would make the process of designing one of performing examples of movement. Designers can design by performing movements that emerge out of their sensorimotor skills without ever having to form explicit, detailed representations of the movements they are performing. Fiebrink et al. [13] have applied interactive machine learning to gestural control of digital musical instruments. Their participants noted that interactive machine learning provided a fundamentally different way of designing which focused on direct embodied movement rather than analyzing gestures in terms of specific features.



Fig. 1. The actor being motion captured (left) and a participant interacting with the virtual character (right)

4 Non-representational Interaction Design for Video Game Characters

This section presents a prototype that attempts to use Interactive Machine Learning to enable the type of non-representational interaction design that we are proposing. It is a system for designing bodily interaction with video game characters. Two people can play the roles of the video game character and the player, showing how the character should respond by acting out the movements themselves. This allows them to design movements in a natural way, by moving, rather than having to think about explicit representations of their movement. The motion of both participants are recorded and synchronized. This data is then used as input to a machine learning algorithm which learns a model for automatically controlling a video game character so that it responds in the same way as the people designing it.

The capture set up enables two people to improvise interaction in order to design the virtual agent. One participant plays the part of the agent while the other plays the part of the player. The two participants are in different spaces and their movements are streamed live to each other. The movements of the performer playing the agent are motion captured using an OptitrackTM optical motion capture system and mapped live onto the agent, which can be seen live by the other participant. This second participant's movements are recorded using a Microsoft KinectTM consumer motion tracking device. His or her movements are visible to the other participant as a live video stream. Both sets of data are synchronised and recorded as input to our machine learning system. For interaction with the AI agent, the participants' set up is identical; they interact with a virtual agent via the Kinect, but in this case the agent's actions are selected by the learnt model rather than being controlled live by a participant in motion capture.

Once data had been recorded the participants use a desktop computer to select clips from the recorded data that will be used as input to train the machine learning classifier. The resulting clips contain two types of data. Firstly they contain a range of motion capture data which can be used to animate the virtual agent in response to the player's action. Secondly they contain kinect data, which

can be used to classify the player's actions. The kinect data at the start of the clip is taken to be the action that produced the response contained in the clip and this is used as input for training the machine learning engine, for which we used a decision tree algorithm.

The final result of our system is a virtual agent that responds live to a player's actions as sensed by the Microsoft KinectTM. The task of the decision tree is therefore to determine which action the participant had performed and then select a suitable response from the agent. The decision tree is used to classify the participant's posture into one of the possible actions. This classification is then used to select an animation clip to play back on the virtual agent.

We tested the system by working with a physical performer to design an interactive character. The performer was a professional who worked in the theatre and also taught performance, and whose practice centred on physical body movement. As such he had considerable expertise in movement and was well suited to designing bodily interaction. He worked with a number of members of the public to design the interaction and was encouraged to use his own working practices as far as possible. He used a long process of physical improvisation, which is a key part of his performance practice. This enabled him to design the detail of the movement. While he was readily able to talk at a high level about the movements and given them names, movement improvisation was key to working out the details of the movements and interactions. The prototype enabled him to design the interaction entirely through this process of physical improvisation without ever having to make explicit low level details of movement. Our non-representation approach therefore fitted easily with his practice as performer and movement expert.

5 Conclusion

Sensorimotor theory provides a powerful lens for understanding both what is compelling about bodily interaction and how it can fail. Seeing bodily interfaces as a way of leveraging our sensorimotor skills that are left out of traditional representational user interfaces helps explain the intuition that this type of interface is some how "natural" while also explaining how many actual examples fail to be natural by falling back on representational gestures. More importantly, it shows us a way forward for how we can create interfaces that do make better use of our sensorimotor skills. Since our sensori-motor skills are highly non-representational, design approaches that rely on representing actions will never capture them appropriately. Interactive machine learning on the other hand allows people to design interactions by performing actions, thus allowing them to directly make use of their non-representational skills. The examples presented in this paper provide an idea of what these new design methods might be like, while the central argument is a call to make sure we do change the way we design interfaces to adapt to our non-representational skills.

References

1. Krueger, M.W.: Responsive environments. In: Proceedings of the National Computer Conference, AFIPS 1977, June 13-16, pp. 423-433. ACM, New York (1977)
2. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., Rasamimanana, N.: Continuous realtime gesture following and recognition. In: Kopp, S., Wachsmuth, I. (eds.) *GW 2009*. LNCS (LNAI), vol. 5934, pp. 73-84. Springer, Heidelberg (2010)
3. Fails, J.A., Olsen Jr., D.R.: Interactive machine learning. In: Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI 2003, pp. 39-45. ACM, New York (2003)
4. Antle, A.N., Corness, G., Droumeva, M.: What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers* 21(1-2), 66-75 (2009)
5. Fiebrink, R., Cook, P.R., Trueman, D.: Human model evaluation in interactive supervised learning. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI 2011, pp. 147-156. ACM, New York (2011)
6. Snibbe, S.S., Raffle, H.S.: Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, pp. 1447-1456. ACM, New York (2009)
7. Fergus, P., Haggerty, J., Taylor, M., Bracegirdle, L.: Towards a whole body sensing platform for healthcare applications. In: England, D. (ed.) *Whole Body Interaction*. Human-Computer Interaction Series, pp. 135-149. Springer, London (2011)
8. Norman, D.A.: Natural user interfaces are not natural. *Interactions* 17(3), 6-10 (2010)
9. Jacob, R.J., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-based interaction: a framework for post-wimp interfaces. In: *CHI 2008: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pp. 201-210. ACM, New York (2008)
10. Dourish, P.: *Where The Action Is: The Foundations of Embodied Interaction*. MIT Press (2001)
11. Talbot, J., Lee, B., Kapoor, A., Tan, D.S.: Ensemblematrix: interactive visualization to support machine learning with multiple classifiers. In: *CHI 2009: Proceedings of the 27th International Conference on Human Factors in Computing Systems*, pp. 1283-1292. ACM, New York (2009)
12. Kulesza, T., Stumpf, S., Wong, W.K., Burnett, M.M., Perona, S., Ko, A., Oberst, I.: Why-oriented end-user debugging of naive bayes text classification. *ACM Trans. Interact.* 1(1), 2:1-2:31 (2011)
13. Fiebrink, R.: *Real-time Human Interaction with Supervised Learning Algorithms for Music Composition and Performance*. PhD thesis, Princeton University, Princeton, NJ, USA (January 2011)